DOCUMENT RESUME

ED 157 088 CS 204 243

AUTHOR Kelso, Genell
TITLE Spelling Correlation on Handwritten vs. Multiple-Choice Tests.
PUB DATE Jun 78
NOTE 60p.; D.Ed. Practicum, Nova University

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS College Freshmen; Educational Research; English Instruction; Higher Education; *Multiple Choice Tests; *Spelling; *Spelling Instruction; *Test Interpretation; Test Items; *Test Validity

ABSTRACT

In order to determine if multiple choice spelling tests are as effective as handwritten tests, 265 college freshmen were tested on the same spelling words by traditional oral dictation and then by five-option multiple choice questions. Results were compared to examine the efficacy of multiple choice testing. Of a possible 20 points, the mean score on the multiple choice test was 10.77; on the written test, 8.78; and of words spelled correctly on both, the mean was 7.07. It was found that the multiple choice test score was inflated in that it provided a higher and less accurate reflection of the student's spontaneous spelling ability than a written spelling test. The multiple choice spelling test, however, does offer the advantages of greater grading speed, accuracy, and objectivity. It can be concluded that alternating written and multiple choice tests throughout a spelling program would be ideal. Each testing method could provide correction and compensation for the other's disadvantages, while maintaining maximum advantages of each. (Author/JF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157088

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SPELLING CORRELATION

ON HANDWRITTEN VS. MULTIPLE-CHOICE TESTS

Applied Educational Research and Evaluation Module

by

Genell Kelso, M. A.

El Centro College
Dallas, Texas

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Genell Kelso

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Dr. Anita G. Barrett
Cluster Coordinator
Dallas I Cluster

A PRACTICUM PRESENTED TO NOVA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF EDUCATION

NOVA UNIVERSITY

June 16, 1978

CS204243

TABLE OF CONTENTS

Introduction 1

Background and significance of the study 1

Procedures

- Population 6
- Definition of terms 6
- Limitations of the study 6
- Hypothesis 7
- Basic assumptions 7
- Procedures for collecting data 8
- Procedures for treating data 12

Results 16

Discussion, Implications, and Recommendations 36

Appendices

- Appendix A - Answer Strip for Written Spelling Test . . . 44
- Appendix B - Multiple-choice Spelling Test 45
- Appendix C - Frequency Distributions of Written Test Responses 46
- Appendix D - Statistical Computations 53

Bibliography 55

LIST OF TABLES

Table 1 - Word selection basis 10

Table 2 - Frequency distribution of raw scores 17

Table 3 - Statistical results of tests 19

Table 4 - Descriptive item-response analysis 21

Table 5 - Difficulty index (proportion passing each item) . . . 26

Table 6 - Order of ease-to-difficulty 27

Table 7 - Discrimination index table 29

Table 8 - Statistical summary of experimental results 32

Table 9 - Scatter diagram of multiple-choice and written test scores' correlation 33

Table 10 - Written test scores predicted from multiple-choice test scores 35

Table 11 - Recommended multiple-choice distracters (for future use) 40

SPELLING CORRELATION
ON HANDWRITTEN VS. MULTIPLE-CHOICE TESTS

INTRODUCTION

Given the same spelling words by traditional oral dictation and then in five-option multiple-choice questions, will students get the same scores on each? And will they get the same words right on each?

Multiple-choice spelling tests are currently used at El Centro College (downtown campus of Dallas County Community College District) on a limited basis, in a few English and Developmental Writing sections, and in the Testing Center. This study has sought to determine the efficacy and validity of such multiple-choice spelling tests, as opposed to written spelling tests presented by traditional oral dictation.

The independent variables were a written spelling test with words dictated, and a multiple-choice spelling test covering the same words. Both tests were administered to the same subjects on the same day. Dependent variables were the number right on the written test, the number right on the multiple-choice test, and the number of the same words right on both tests.

BACKGROUND AND SIGNIFICANCE OF THE STUDY

For many community college students, spelling is a serious problem. John A. Higgins, who analyzed sample placement themes written by incoming freshmen students under New York City's Open Admissions program, found that the single most common fault was misspelling, which occurred in 83% of all papers and recurred (i.e. appeared two or more times) in 64%.[1]

[1] John A. Higgins, "Remedial Students' Needs vs. Emphases in Text-Workbooks," College Composition and Communication, 24:188-190 (May, 1973).

And Mina Shaughnessy reaches these conclusions about college-level spelling:

> Of all the encoding skills, spelling tends to be viewed by teachers and students alike as the most arbitrary, the most resistant to instruction, and the least related to intelligence (a myth that has comforted many bad spellers). It is the one area of writing where English teachers themselves will admit ineptness. Outside the academy, however, the response to misspelling is less obliging. Indeed, the ability to spell is viewed by many as one of the marks of the educated person, and the failure of a college graduate to meet that minimal standard of advanced literacy is cause to question the quality of his education or even his intelligence....There are many more "opportunities" for misspelling than for grammatical errors. Every word is a potential misspelling, whereas only certain words in each sentence can figure in grammatical errors.[2]

Palmer reports that various methods have been used to test spelling:

> Some try to put their finger on the exact error in spelling, asking the student to correct a word by deleting a consonant or a vowel, by adding a consonant or a vowel, by transposing. Most of these are more complicated than they are worth; objective testing of spelling is just not very satisfactory. Perhaps the simplest procedure is to present the student with a group of four words, among which none, all, or some of which are misspelled.[3]

Although a review of the literature reveals numerous experiments on spelling teaching techniques (mostly at elementary levels), as well as numerous experiments comparing multiple-choice with other forms of testing, none located to date have investigated the relative effectiveness of multiple-choice testing of spelling. This area, however, bears investigation for the following reasons:

Both handwritten and multiple-choice spelling tests would seem to have certain inherent advantages (as well as disadvantages, to be

[2]Mina P. Shaughnessy, Errors and Expectations: A Guide for the Teacher of Basic Writing (New York: Oxford University Press, 1977), pp. 161-162.

[3]Osmond E. Palmer, "Evaluation of Communication Skills," Evaluation in Higher Education, Paul L. Dressel and others, ed. Herold C. Hunt (Boston: Houghton Mifflin Company, 1961), p. 215.

discussed below). Multiple-choice testing offers speed (including machine scoring) and objectivity. For example, the subjective grading factors of handwriting are avoided ("Is this an a or an o?").

> When one reads published reports on misspelling, the printing of the report obscures a great truth -- that many words in student writing have been formed so carelessly that one cannot be sure whether or not they are misspelled, or, if so, how. More use of photographs of handwritten words in published research would help to convey some sense of the problem the spelling investigator faces in illegibility.[4]

A multiple-choice test, by avoiding illegibility problems, offers greater objectivity.

> One characteristic of a test that influences its reliability is its objectivity -- that is, the extent to which different scorers can, by prescribed procedures, arrive at the same scores for the individuals tested. It is evident from this definition that objectivity in this sense sets an upper bound for a test's reliability. Multiple-choice and true-false tests are usually completely objective....Completion tests requiring a brief written answer and essay tests involve some judgment in scoring and hence have less than perfect objectivity.[5]

Further comment on the advantages of multiple-choice testing is offered by Wick:

> A glance at the major published tests of achievement and aptitude will illustrate that the multiple-choice item is the basic building block of this industry. Surely nine of ten items in such tests will be in a multiple-choice format. The reasons for the high multiple-choice emphasis include ease of scoring, higher discrimination power per item, and the ability to measure all cognitive levels with a single item-type. The scoring issue is probably the critical one.[6]

But as Wick then goes on to point out, "What's good for the testing

[4]Richard Braddock, Richard Lloyd-Jones, and Lowel Schoer, Research in Written Composition (Champaign, Illinois: National Council of Teachers of English, 1963), p. 50.

[5]Joe L. Saupe, "Technical Considerations in Measurement," Evaluation in Higher Education, Paul L. Dressel and others, ed. Herold C. Hunt (Boston: Houghton Mifflin Company, 1961), p. 443.

[6]John W. Wick, Educational Measurement: Where Are We Going and How Will We Know When We Get There? (Columbus, Ohio: Charles E. Merrill Publishing Company, 1973), p. 123.

industry may not be good for the classroom teacher."[7]

Multiple-choice testing of spelling thus would seem to offer worthwhile advantages if such testing can be shown to be valid and comparable to the student's written spellings.

The major disadvantage of multiple-choice testing -- that the student recognizes the answer rather than recalling/supplying/constructing it -- and the ramifications of this disadvantage are widely noted in the literature, as this sampling represents:

> Most of the items appearing in commercially published standardized tests are not supply-type. True-false, matching, and multiple-choice items have a restricted number of response possibilities. Rather than recall a skill, some information, or a thought, as the student must do with a supply-type item, the task is to recognize or select the most correct answer from a limited number of possibilities. The student is required to discriminate from among a list of plausible responses the one which is most correct.[8]

> Ability to select a correct response among several alternatives is not a sure indication of ability to recall without cue.[9]

> One may not infer that the proportion of a group which passes an item is identical with the proportion possessing "the knowledge" required by it. In multiple-choice questions, particularly, this is a most tenuous inference. The nature of the incorrect alternatives often has a great influence on the attractiveness of the correct one. An item analysis occasionally demonstrates that presumed wrong answers may occasionally be more attractive to the ablest students than is the intended correct answer. The phenomenon of guessing may also be assumed to cast doubt on interpretation of difficulty values in any absolute manner. Chance success by the ignorant will occur in a few cases. Even the misinformed may select a right answer for a wrong reason.[10]

[7] Ibid.

[8] Ibid., p. 109.

[9] Paul L. Dressel, Handbook of Academic Evaluation: Assessing Institutional Effectiveness, Student Progress, and Professional Performance for Decision Making in Higher Education (San Francisco: Jossey-Bass Publishers, 1976), p. 213.

[10] Saupe, pp. 450-451.

> The use of objective tests of attainment, and particularly the multiple-choice item format, increases year by year in our school system. Whatever psychometric advantages such tests exhibit in a research context,...the feature of a multiple-choice test item that arouses the greatest hostility among teachers and examiners is that it requires recognition of, rather than the construction of, the correct response. This has implications for both the validity and the reliability of the test; the validity, because the test constructor is usually more interested in assessing the students' ability to solve a problem than his capacity to recognize the solution when it is shown to him; the reliability, because the student has a very real chance (a fifth, a quarter, or sometimes even a half) of guessing the right answer, even when he has absolutely no idea how to solve the problem.[11]

The desired skill and ultimate objective of any spelling instruction, however, is that the student be able to spell spontaneously the words in question -- handwritten (or at least unassisted by visual options to choose amongst). The present study has sought to determine the efficacy and validity of multiple-choice tests currently in use at El Centro. Serious discrepancies between written and multiple-choice spelling scores -- if the null hypothesis must be rejected in the event, for example, that students make significantly higher and/or uncorrelated scores on the multiple-choice test -- would indicate that multiple-choice spelling tests are of questionable value, and may not reflect the students' true spelling abilities. But high correlation between the longhand and multiple-choice scores -- if the null hypothesis stands -- would support the use of multiple-choice spelling tests as an efficient method of achieving comparable testing results.

[11]B. Choppin, "Guessing the Answer on Objective Tests," British Journal of Educational Psychology, 45:206 (June, 1975).

PROCEDURES

POPULATION

Subjects were all students enrolled in El Centro's English 101 day sections, Fall 1977, whose instructors agreed to participate in the study, and all who reported for class on the day tests were administered: N = 265.

DEFINITION OF TERMS

1. Spelling words - words presented were similar to a pretest, in that students had not previously studied for this test. Words chosen did not include homonyms/homophones (i.e. principal/principle) but only words with a single correct spelling. Derivation of the word list will be explained below.
2. Handwritten spelling - words as written ("longhand" or "cursive" or printed) by students in response to teacher's oral dictation. This presumably would equate with the student's spelling in his own compositions.
3. Multiple-choice test - five-option questions answered by the student's circling his choice. These tests were then keypunched to expedite multivariate analysis by computer.

LIMITATIONS OF THE STUDY

From the scorer's standpoint, illegibility is the primary limitation of written spelling. In the present study, the scorer has interpreted (insofar as possible) the letters comprising each spelling word written by students. When such interpretation was not possible, or when only a partial word was entered, that word was disallowed as though unanswered, and tabulated as "no response."

HYPOTHESIS

H_o: $\bar{X}_{mc} = \bar{X}_w$

Students' mean score of words spelled correctly on the multiple-choice test will equal students' mean score of words spelled correctly on the written test. Acceptance of this null hypothesis would reflect the feasibility of using the multiple-choice method to test spelling, as an adequate representation of students' written spelling. Rejection of this null hypothesis would reflect that discrepancies between written and multiple-choice testing of spelling are too great to consider the multiple-choice method as an adequate representation of students' written spelling.

BASIC ASSUMPTIONS

It is assumed that the total group of students involved in the experiment involves a normal distribution of previous spelling experience, ability, and training, and that good spellers and poor spellers are randomly distributed through this student population.

It is assumed that a student who "knows" how to spell a given word will spell it in the same correct way every time.

It is assumed that a student who spells the same given word one way one time and another way minutes later (even though correctly spelled once) may do so by carelessness, but more likely, does not "know" how to spell that word; instead, he is unsure and (at least on some occasions) guessing.

It is assumed that all words included are familiar (at least in reading-recognition) to all student-subjects.

It is assumed that, as an inner-city community college, El Centro's student ability level may be somewhat below average.

PROCEDURES FOR COLLECTING DATA

1. A list of words commonly misspelled by freshmen was derived for inclusion on the two spelling tests. For this, two sources were used and cross-matched: The Pollock list, and Kelso's spelling study conducted at El Centro.

   The Seventh Mental Measurements Yearbook recommended only one of the spelling tests it reviewed: The McGraw-Hill Basic Skills System Spelling Test for Grades 11-14.

   > The words selected for these tests were drawn from Pollack's (sic) study of misspellings and seven other lists of words commonly misspelled by high school and college students....The test's words were selected from a frequency count of 4,482 different misspelled words and word groups of students that Pollack (sic) collected from 599 college English teachers in 1950.[12]

   In other research, Dean Thomas Clark Pollock studied 31,375 misspellings in the written work of college students. He identified the hundred words most frequently misspelled (more than 43 times), including nineteen designated as most misspelled of all, then the next 550 words most frequently misspelled.[13]

   During Fall semester 1976, a pretest-posttest spelling experiment was conducted at El Centro by Kelso; one group of students studied spelling rules by classroom drill while another studied by self-instructional package. This study, which utilized multiple-choice

[12]Oscar Krisen Buros, ed., "Spelling," The Seventh Mental Measurements Yearbook (Highland Park, New Jersey: The Gryphon Press, 1972), 1:512-513. Quoting Thomas Clark Pollack (sic), "Spelling Report," College English, 16:102-109 (November, 1954).

[13]Thomas Clark Pollock and William D. Baker, The University Spelling Book (Englewood Cliffs, New Jersey, 1955). Quoted in John C. Hodges and Mary E. Whitten, Harbrace College Handbook (New York: Harcourt Brace Jovanovich, Inc., 1972), pp. 176-177.

testing, involved a total of 2357 misspellings.[14]

From these two lists, a pool of commonly misspelled words within both testing methods was compiled. Words included were not necessarily those misspelled most often (as shown in Table 1, p. 10), but were selected among these to give a good variety of common error types. For the present study, words chosen excluded homonyms/homophones (i.e. principal/principle, there/their/they're, etc.); only words with a single correct spelling were chosen.

2. A test length of twenty items was chosen, based on the following rationale: Lord advises that efficiency involves three factors: number of items, length of items, and ability level.

> At high ability levels,.there is little random guessing and the relative efficiency of...tests is determined by their length.... At low ability levels the effect of random guessing becomes of overwhelming importance. At an ability level of 215, the rank order of test efficiency is precisely the opposite of the order at 800. Thus, for any pair of tests, one test is better than the other over a certain range of ability, but worse over the complementary range of ability....The effect of decreasing the number of choices per item while lengthening the test proportionately is to increase the efficiency of the test for high-level examinees and to decrease its efficiency for low-level examinees.[15]

This implied that, for students at El Centro College (assumed to be somewhat below average ability level), greater efficiency could be achieved using fewer items with more choices per item. Therefore five-option questions were written for the multiple-choice portion of the test.

[14]Genell Kelso, Spelling Improvement Units Taught by Classroom Drill vs. Self-Instructional Package: Their Development and Comparative Evaluation (Ft. Lauderdale, Florida, Nova University), September 1977 (ERIC, number not yet available), pp. 92-93.

[15]Frederic M. Lord, "Optimal Number of Choicer Per Item -- A Comparison of Four Approaches," Journal of Educational Measurement, 14:36 (Spring, 1977).

TABLE 1

WORD SELECTION BASIS

The following twenty words were selected:	In Pollock's study of written misspellings	In Kelso's study of multiple-choice misspellings
ACCIDENTALLY		Missed by 18 of 55 students
BELIEVE	Among 19 most misspelled	
CATEGORY	Among 100 most misspelled	Missed by 18 of 55 students
CONSCIENCE		Missed by 40 of 55 students
DISAPPOINT		Missed by 23 of 55 students
DEFINITELY	Among 100 most misspelled	Missed by 15 of 55 students
ENVIRONMENT	Among 100 most misspelled	Missed by 17 of 55 students
EXISTENCE	Among 19 most misspelled	Missed by 23 of 55 students
IRRELEVANT		Missed by 34 of 55 students
MISSPELLED		Missed by 32 of 55 students
OCCASIONALLY	Among 19 most misspelled	Missed by 24 of 55 students
OCCURRING	Among 100 most misspelled	Missed by 45 of 55 students
PRECEDE	Among 19 most misspelled	
PRIVILEGE	Among 19 most misspelled	Missed by 22 of 55 students
RECEIVING	Among 19 most misspelled	
REFERRED	Among 19 most misspelled	Missed by 27 of 55 students
SEPARATE	Among 19 most misspelled	Missed by 15 of 55 students
SIMILAR	Among 19 most misspelled	Missed by 15 of 55 students
UNNECESSARY	Among 100 most misspelled	
WRITING	Among 100 most misspelled	Missed by 18 of 55 students

Similar advice comes from Saupe, saying that a "characteristic of tests which affects their reliability is length....If the evaluator is interested only in the average performance of a large group, he may obtain satisfactory results by using a short test even though it will produce relatively unreliable individual scores."[16]

A twenty-item test was chosen partly to provide a maximum student sample population, since it was believed that more teachers would be willing to participate in the study if the test were relatively short and required only a small amount of class time. (This proved successful: teachers of sixteen of the eighteen sections agreed to participate.)

3. The two test forms were written, to be administered by procedures similar to the alternate-forms method. Although this term is usually applied to a pair of tests using different items in the same method (e.g. both multiple-choice, or both true-false), its procedures have been applicable to the present study.

> With the alternate-forms method two equivalent forms of a test are made up and a group of pupils is given both forms in order that their scores on each form can be compared. If the two sets of scores show a high degree of relationship, or a high correlation, the test is said to be reliable. Obviously the higher the correlation, the more reliable the test will be.
> The alternate-forms method gives a coefficient of equivalency if both forms of the test are administered to the same group of pupils at the same time, since the purpose of the method is to develop two forms on which pupils obtain equivalent scores and equivalent rank within their groups.[17]

Anastasi defines alternate-form reliability: "The correlation between the scores obtained on the two forms represents the

[16]Saupe, pp. 443-444.

[17]John A. Green, Teacher-Made Tests (New York: Harper & Row, Publishers, 1963), p. 94.

reliability coefficient of the test. If the two forms are administered in immediate succession, the resulting correlation shows reliability across forms only, not across occasions."[18]

4. Those teaching English 101 day sections during Fall 1977 were asked to participate, and a testing schedule was set up.
5. A cassette recording of the oral dictation was made for use in all sections, to insure identical pronunciations, timing, context sentences, etc.
6. Teachers were given test materials for their classes, with the following instructions:

    1. Approximately 15 minutes before end of period, give each student a long answer strip, and ask him to enter name, social security number, and section.
    2. When they've completed the heading, turn on tape, which reads instructions at top of strip, then dictates 20 words. Tape runs 6½ minutes.
    3. Collect answer strips, and distribute multiple-choice test. Ask them to enter heading <u>first</u>, then begin circling answers. They'll turn in multiple-choice test as they finish and leave.

(Answer strip for written spelling test, and a copy of the multiple-choice spelling test, are included as Appendices A and B, respectively.)

7. Both tests were scored manually, then keypunched for computer analysis.

PROCEDURES FOR TREATING DATA

As advised by Campbell and Stanley, the present study will include several dependent variables:

> Experiments may be multivariate in either or both of two senses. More than one "independent" variable (sex, school grade, method of teaching arithmetic, style of printing type, size of printing type, etc.) may be incorporated into the design and/or more than one

[18]Anne Anastasi, <u>Psychological Testing</u> (New York: MacMillan Publishing Co., Inc., 1968), pp. 80-81.

"dependent" variable (number of errors, speed, number right, various tests) may be employed.[19]

1. For each student, the following information was collected/tabulated: score on written spelling test, score on multiple-choice spelling test, and number of words correctly spelled on both tests. Frequency distributions were tabulated (Table 2)
2. Mean scores were calculated for the group: on written spelling test, on multiple-choice spelling test, and on number of words correctly spelled on both tests.
3. In addition to these mean scores, the following information was calculated: maximum score, minimum score, range, median, mode, mean deviation, standard deviation, and variance (Table 3).
4. The Kuder-Richardson Formula 21 was applied to measure reliability of both written and multiple-choice tests, using the following formula: reliability = $1 - \frac{M(n - M)}{ns^2}$

   "The reliability of the test depends on just three quantities: the number of items, the standard deviation, and the mean (average). Use n for number of items (not number of students, remember!), s for the standard deviation, and M for the mean."[20]
5. Descriptive item analysis was performed on each word: On the multiple-choice test, the following procedure as described by Saupe was used:

[19]Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally College Publishing Company, 1963), p. 4.

[20]Paul B. Diederich, "Short-Cut Statistics for Teacher-Made Tests," Statistics and Measurement in the Classroom, ed. Carl F. Hereford, Luiz F. S. Natalicio, and Susan J. McFarland (Dubuque, Iowa: Kendall/Hunt Publishing Company, 1969), p. 150.

With multiple-choice items, item analysis procedures may be extended to include a study of responses to incorrect answers as well as a determination of difficulty and discrimination based on the right answer....For example, if, in the simple procedure for studying discrimination, responses to all alternatives were tallied, the result for a four-choice item might appear as follows:

	Number selecting:			
	Choice A	Choice B	Choice C	Choice D
High criterion group (20)	3	12	0	5
Low criterion group (20)	11	6	1	2

Choice B is the correct answer. The simple discrimination index for this item would be .30 and the estimate of its difficulty would be .45. [21]

The entire group was included (high + low criterion groups). The criterion used was the mean score, which divided the group into two approximately equal halves.

On the written test, each item was analyzed by the following procedure, similar to that above: The same possible answers of the multiple-choice test were named, plus an additional category, "OTHER." Responses were thus tallied (see Table 4), and "OTHER" responses were listed/tabulated thereafter (Appendix C).

6. The difficulty of each item was calculated, using the formula,

$$\frac{\text{number right}}{\text{number answering}}.$$

The two basic characteristics of items usually investigated in item analyses are <u>difficulty</u> and <u>discrimination</u>. The difficulty of an item is reflected by a summary of the responses of the group to it. If the items are scored as either correct or incorrect, the proportion of the group passing the item is usually taken as an index of its difficulty....defined as the proportion of individuals passing an item, may range from 0 for an item which no one passes to 1.00 for an item passed by everyone.[22]

7. The simple discrimination index for each item was calculated, using

[21] Saupe, pp. 454-455.

[22] <u>Ibid</u>., p. 450.

the following formula:

High group proportion right - low group proportion right = discrimination

8. The null hypothesis was tested, using independent t-test, two-tailed analysis at the 0.01 level of significance:

H_0: $\bar{X}_{mc} = \bar{X}_w$ Students' mean score of words spelled correctly on the multiple-choice test will equal students' mean score of words spelled correctly on the written test.

The alternate to this hypothesis is H_a: $\bar{X}_{mc} \neq \bar{X}_w$.

The rationale was as follows:

Acceptance of this null hypothesis would reflect the feasibility of using the multiple-choice method to test spelling, as an adequate representation of students' written spelling.

Rejection of this null hypothesis and acceptance of the alternate hypothesis would reflect that discrepancies between written and multiple-choice testing of spelling are too great to consider the multiple-choice method as an adequate representation of students' written spelling.

Because multiple-choice testing offers advantages of speed, accuracy, and objectivity (as discussed on p. 3, above), retaining these advantages would be desirable. In testing the null hypothesis, the acceptance region was therefore set to be as large as practicable, to minimize the chance of a Type I error (rejecting the null hypothesis unnecessarily). Because the risk of a Type I error would have been more serious than the risk of a Type II error, the 0.01 level of significance was used in testing this hypothesis.

9. A scatter diagram was plotted, to analyze correlation between

students' multiple-choice and written scores.

10. The strength of linear correlation was calculated, using Pearson's coefficient formula:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\ \sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

in which n = number of students taking the pair of tests (265); x = individual scores on multiple-choice test; and y = individual scores on written test.

11. The estimated regression line was then calculated by the least squares method, using the following formula:

$$Y \text{ predicted} = m_y + b(x - m_x)$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Thereafter, a specific formula and table were derived for predicting a student's score on the written spelling test (given his score on the multiple-choice test).

RESULTS

1. Each student received three scores: score on written spelling test, score on multiple-choice spelling test, and number of words correctly spelled on both tests. Frequency distributions are shown in Table 2.

2. Each student's scores were then compared, to determine how many had done equally well on both versions of the test -- vs. scoring higher on one or the other:

175 (66%) had higher score on multiple-choice than written test.

51 (19%) had higher score on written than multiple-choice test.

34 (13%) had same score on written and multiple-choice tests, but a lower number correct on <u>both</u> tests.

5 (2%) had same score on multiple choice, written test, and correct on both.

265

TABLE 2

FREQUENCY DISTRIBUTION OF RAW SCORES

Raw score	Multiple choice	Written test	Correct on both tests
20	1	-	-
19	3	5	3
18	6	3	2
17	10	10	4
16	5	9	5
15	22	12	8
14	19	16	6
13	23	19	10
12	27	13	19
11	26	10	16
10	17	14	13
9	28 (mode)	18	12
8	21	18	10
7	25	20	22
6	12	22 (mode)	16
5	15	19	24
4	3	18	20
3	-	12	25 (mode)
2	1	11	19
1	1	8	13
0	-	8	18
	N = 265	265	265

3. From individual scores, group statistics were computed: maximum score, minimum score, range, mean, median, mode, mean deviation, standard deviation, and variance (Table 3).

   As is apparent in these raw score tabulations, students achieved the highest mean score on the multiple-choice test: 10.77. The mean score on the written test was almost exactly two points lower: 8.78. The lowest mean score was the number of words students got right on both tests: 7.07. This last mean score reflects students' frequent spelling of the same word one way on the written test and a different way on the multiple-choice test (implying uncertainty and guessing.)

4. Reliability of each test was calculated, using Kuder-Richardson Formula 21, in which n = number of test items, M = mean, and s = standard deviation: $\text{reliability} = 1 - \frac{M(n - M)}{ns^2}$

   $\text{Multiple choice test's reliability} = 1 - \frac{10.77\ (20 - 10.77)}{20\ (13.34)} = .627$

   $\text{Written test's reliability} = 1 - \frac{8.78\ (20 - 8.78)}{20\ (23.88)} = .794$

   The reliability of each test falls within acceptable range. Within this range, however, the reliability of the multiple-choice test (on which the mean score was 10.77) is relatively low: .627.

   But the reliability of the written test is comparatively high: .794. (On the written test, the lower mean score was 8.78.)

TABLE 3

STATISTICAL RESULTS OF TESTS

N = 265	Multiple choice	Written test	Correct on both tests
Maximum score	20	19	19
Minimum score	2	0	0
Range	18	19	19
Mean	10.77	8.78	7.07
Median	11	8	6
Mode	9	6	3
Mean deviation	3.04	4.15	4.03
Standard deviation	3.65	4.89	4.79
Variance (s^2)	13.34	23.88	22.96

5. Descriptive item analysis was performed on each word. For comparison purposes, responses on the multiple choice and written tests have been compiled on a single analysis (Table 4).

For each word of the multiple-choice test, responses to each of the five options were tabulated. These same five options were also used to classify responses on the written test, with the addition of a sixth category, "other responses." For both tests, a final "no response" column was tabulated.

The population taking each test was then subdivided into high and low criterion groups (to begin discrimination analysis). The criterion used was the mean score, which divided each population into two approximately equal halves:

On the multiple-choice test, the mean was 10.77.

140 scored above the mean (11 or higher);

125 scored below the mean (10 or lower).

On the written test, the mean was 8.78.

129 scored above the mean (9 or higher);

136 scored below the mean (8 or lower).

(Note: 111 students scored in the high group on both tests.

29 scored in high group on multiple-choice but low group on written test.

18 scored in high group on written test but low group on multiple-choice.

125 scored in low group on both tests.)

For each test option, responses were subdivided into those made by high (above mean) and low (below mean) groups.

TABLE 4

DESCRIPTIVE ITEM-RESPONSE ANALYSIS

ITEM 1 I -- you.

	BELEIF	BELIEF	BELIEVE	BELEIVE	BELEEVE	Other* Response	No Response
MC total	1 (.4%)	4 (1.5%)	233(87.9%)	25 (9.4%)	1 (.4%)		1 (.4%)
MC high	-	1	136	3	-		-
MC low	1	3	97	22	1		1
Wr total	1 (.4%)	2 (.8%)	212(80.%)	16 (6.%)	1 (.4%)	33(12.5)	-
Wr high	1	1	120	5	-	2	-
Wr low	-	1	92	11	1	31	-

ITEM 2 They have -- offices.

	SEPERETE	SEPARETE	SEPERATE	SEPARATE	SEPPERATE	Other Response	No Response
MC total	-	7 (2.6)	90 (34.)	165(62.3)	1		2 (.8)
MC high	-	1	27	112	-		-
MC low	-	6	63	53	1		2
Wr total	-	-	75(28.3)	144(54.3)	-	43(16.2)	3(1.1)
Wr high	-	-	24	102	-	3	-
Wr low	-	-	51	42	-	40	3

ITEM 3 Their ideas are --.

	SIMALAR	SIMILAR	SIMULAR	SIMALIAR	SIMILIAR	Other Response	No Response
MC total	14(5.3)	177(66.8)	35(13.2)	5(1.9)	34(12.8)		-
MC high	2	109	5	2	22		-
MC low	12	68	30	3	12		-
Wr total	6(2.3)	156(58.9)	24(9.1)	2 (.8)	35(13.2)	36(13.6)	6(2.3)
Wr high	1	102	4	-	20	2	-
Wr low	5	54	20	2	15	34	6

ITEM 4 He is in amateur --.

	CATAGORY	CATEGORY	CATOGORY	CATAGARY	CATEGARY	Other Response	No Response
MC total	98(37.)	146(55.1)	17(6.4)	4(1.5)	-		-
MC high	36	100	3	1	-		-
MC low	62	46	14	3	-		-
Wr total	80(30.2)	101(38.1)	8(3.)	-	1(.4)	69(26.)	6(2.3)
Wr high	39	74	4	-	-	12	-
Wr low	41	27	4	-	1	57	6

ITEM 5 She is -- a book.

	WRITING	WRITEING	WRITTING	WRITTEING	RITEING	Other Response	No Response
MC total	196(74.)	8(3.)	59(22.3)	1(.4)	1(.4)		-
MC high	127	1	12	-	-		-
MC low	69	7	47	1	1		-
Wr total	197(74.3)	4(1.5)	50(18.9)	-	-	13(4.9)	1(.4)
Wr high	123	-	6	-	-	-	-
Wr low	74	4	44	-	-	13	1

* "Other responses" are tabulated in Appendix C.

(Table 4, continued)

ITEM 6 He -- to the dictionary.

	REFERED	REFFERED	REFFERD	REFERRED	REFFERRED	Other Response	No Response
MC total	117(44.2)	42(15.8)	9(3.4)	93(35.1)	4(1.5)		-
MC high	46	20	3	69	2		-
MC low	71	22	6	24	2		-
Wr total	73(27.5)	14(5.3)	2(.8)	135(50.9)	2(.8)	33(12.5)	6(2.3)
Wr high	22	4	-	98	1	4	-
Wr low	51	10	2	37	1	29	6

ITEM 7 The eagle's -- is threatened.

	EXISTANCE	EXISTANSE	EXISTENCE	EXISTENSE	EXSISTENCE	Other Response	No Response
MC total	130(49.1)	7(2.6)	107(40.4)	4(1.5)	14(5.3)		3(1.1)
MC high	59	-	73	2	4		2
MC low	71	7	34	2	10		1
Wr total	84(31.7)	-	77(29.1)	2(.8)	4(1.5)	82(30.9)	16(6.)
Wr high	42	-	65	1	2	18	1
Wr low	42	-	12	1	2	64	15

ITEM 8 He -- smokes cigars.

	OCASIONALY	OCASSIONALY	OCCASSIONALLY	OCCASIONALLY	OCCASSIONALLY	Other Resp.	No Resp.
MC total	8(3.)	35(13.2)	51(19.2)	144(54.3)	24(9.1)		3(1.1)
MC high	2	7	22	91	18		-
MC low	6	28	29	53	6		3
Wr total	3(1.1)	6(2.3)	14(5.3)	104(39.2)	27(10.2)	94(35.5)	17(6.4)
Wr high	-	1	9	76	23	20	-
Wr low	3	5	5	28	4	74	17

ITEM 9 Voting is a --.

	PRIVILEDGE	PRIVELEDGE	PRIVILEGE	PRIVELEGE	PRIVILAGE	Other Response	No Response
MC total	69(26.)	17(6.4)	131(49.4)	12(4.5)	35(13.2)		1(.4)
MC high	37	4	90	3	6		-
MC low	32	13	41	9	29		1
Wr total	38(14.3)	10(3.8)	64(24.2)	10(3.8)	25(9.4)	104(39.2)	14(5.3)
Wr high	24	4	56	6	12	27	-
Wr low	14	6	8	4	13	77	14

ITEM 10 Don't pollute the --.

	ENVIORIMENT	ENVIRONMENT	ENVIORMENT	ENVIRAMENT	ENVIORAMENT	Other Response	No Response
MC total	50(18.9)	162(61.1)	44(16.6)	3(1.1)	4(1.5)		2(.8)
MC high	14	115	9	-	2		-
MC low	36	47	35	3	2		2
Wr total	13(4.9)	110(41.5)	15(5.7)	1(.4)	1(.4)	119(44.9)	6(2.3)
Wr high	9	90	2	-	-	28	-
Wr low	4	20	13	1	1	91	6

ITEM 11 A storm is --.

	OCURING	OCCURING	OCCERING	OCURRING	OCCURRING	Other Response	No Response
MC total	1(.4)	183(69.1)	6(2.3)	15(5.7)	56(21.1)		4(1.5)
MC high	-	89	-	6	42		3
MC low	1	94	6	9	14		1
Wr total	5(1.9)	124(46.8)	1(.4)	9(3.4)	76(28.7)	42(15.8)	8(3.)
Wr high	1	62	-	5	57	4	-
Wr low	4	62	1	4	19	38	8

(Table 4, continued)

ITEM 12 She has a guilty --.

	CONSCIUS	CONSCIENCE	CONSCIOUS	CONSCEINCE	CONSCIENSE	Other Response	No Response
MC total	5(1.9)	136(51.3)	101(38.1)	16(6.)	4(1.5)		3(1.1)
MC high	1	88	32	6	2		-
MC low	4	48	69	10	2		3
Wr total	-	98(37.)	45(17.)	2(.8)	1(.4)	96(36.2)	23(8.7)
Wr high	-	79	29	2	-	18	1
Wr low	-	19	16	-	1	78	22

ITEM 13 He -- broke the glass.

	ACCIDENTLY	ACCIDENTALY	ACCIDENTALLY	ACCIDENTILY	ACCIDENTILLY	Other Resp.	No Resp.
MC total	91(34.3)	39(14.7)	122(46.)	6(2.3)	3(1.1)		4(1.5)
MC high	36	9	91	1	1		2
MC low	55	30	31	5	2		2
Wr total	81(30.6)	19(7.2)	105(39.6)	1(.4)	2(.8)	52(19.6)	5(1.9)
Wr high	25	5	93	-	1	5	-
Wr low	56	14	12	1	1	47	5

ITEM 14 June will -- July.

	PRECEDE	PRECEED	PRECEEDE	PROCEDE	PROCEEDE	Other Response	No Response
MC total	127(47.9)	101(38.1)	10(3.8)	22(8.3)	5(1.9)		-
MC high	94	39	1	4	2		-
MC low	33	62	9	18	3		-
Wr total	103(38.9)	86(32.5)	7(2.6)	-	1(.4)	63(23.8)	5(1.9)
Wr high	78	36	2	-	-	13	-
Wr low	25	50	5	-	1	50	5

ITEM 15 She is -- many letters.

	RECIEVING	RECIEVEING	RECEIVING	RECEIVEING	RECEVING	Other Response	No Response
MC total	108(40.8)	15(5.7)	132(49.8)	3(1.1)	6(2.3)		1(.4)
MC high	42	3	92	-	2		1
MC low	66	12	40	3	4		-
Wr total	89(33.6)	2(.8)	127(47.9)	3(1.1)	4(1.5)	37(14.)	3(1.1)
Wr high	38	-	88	2	-	1	-
Wr low	51	2	39	1	4	36	3

ITEM 16 I'm sorry to -- you.

	DISAPPOINT	DISAPOINT	DISSAPOINT	DISSAPPOINT	DISSAPPOINTE	Other Response	No Response
MC total	147(55.5)	53(20.)	39(14.7)	24(9.1)	2(.8)		-
MC high	96	12	17	15	-		-
MC low	51	41	22	9	2		-
Wr total	125(47.2)	41(15.5)	44(16.6)	11(4.2)	1(.4)	41(15.5)	2(.8)
Wr high	86	14	20	8	-	1	-
Wr low	39	27	24	3	1	40	2

ITEM 17 That information is --.

	IRRELEVENT	IRRELEVANT	IRRELAVENT	IRREVELANT	IRREVELENT	Other Response	No Response
MC total	45(17.)	144(54.3)	36(13.6)	24(9.1)	12(4.5)		4(1.5)
MC high	22	90	12	11	3		2
MC low	23	54	24	13	9		2
Wr total	16(6.)	86(32.5)	11(4.2)	9(3.4)	9(3.4)	113(42.6)	21(7.9)
Wr high	7	67	7	7	5	34	2
Wr low	9	19	4	2	4	79	19

(Table 4, continued)

ITEM 18 He -- four words.

	MISS-SPELLED	MISPELED	MISSPELED	MISPELLED	MISSPELLED	Other Response	No Response
MC total	30(11.3)	9(3.4)	19(7.2)	86(32.5)	119(44.9)		2(.8)
MC high	10	1	2	38	87		2
MC low	20	8	17	48	32		-
Wr total	6(2.3)	4(1.5)	-	49(18.5)	128(48.3)	76(28.7)	2(.8)
Wr high	3	2	-	25	78	21	-
Wr low	3	2	-	24	50	55	2

ITEM 19 This trip is --.

	UNNESARY	UNNESISARY	UNECISSARY	UNNECESARY	UNNECESSARY	Other Response	No Response
MC total	1(.4)	9(3.4)	8(3.0)	46(17.4)	200(75.5)		1(.4)
MC high	-	-	-	15	123		1
MC low	1	9	8	31	77		-
Wr total	-	-	-	5(1.9)	106(40.)	147(55.5)	7(2.6)
Wr high	-	-	-	3	85	41	-
Wr low	-	-	-	2	21	106	7

ITEM 20 You are -- right.

	DEFINATLY	DEFINATELY	DEFINETELY	DEFINITELY	DEFINITLY	Response	Response
MC total	25(9.4)	65(24.5)	27(10.2)	117(44.2)	29(10.9)		2(.8)
MC high	5	25	12	86	11		1
MC low	20	40	15	31	18		1
Wr total	18(6.8)	31(11.7)	5(1.9)	72(27.2)	10(3.8)	119(44.9)	10(3.8)
Wr high	8	19	4	65	4	29	-
Wr low	10	12	1	7	6	90	10

6. Difficulty of each item was calculated, using the formula $\frac{\text{number right}}{\text{number answering}}$ to determine the proportion of individuals passing each item.

Comparative difficulty figures of the two tests are shown in Table 5. On 17 of the 20 word-items, a higher proportion passed on the multiple-choice test (by a mean difference of .133, ranging from .003 multiple-choice higher on WRITING to .355 multiple-choice higher on UNNECESSARY). On only three words (REFERRED, OCCURRING, and MISSPELLED -- all dealing with doubling of consonants) was a higher proportion correct on the written test, by a mean difference of .089. A general difficulty ratio for each test was then calculated: the probability of a student's correctly answering any particular question of that quiz. This was calculated by totaling the difficulty indices of its twenty questions (that sum equaling the test's mean score), then dividing by twenty.

The multiple-choice test's difficulty ratio was .538 (a student had slightly better than 50-50 chance of answering any particular question correctly).

The written test's difficulty ratio was .439 (a student had less than 50-50 chance of answering any particular question correctly).

Words were then classified in difficulty-rank order for each test (Table 6). Of five words in the easiest quartile of each quiz, four appeared in the easiest quartile of <u>both</u>: BELIEVE, WRITING, SIMILAR, and SEPARATE. Three words appeared in the most difficult quartile of both: DEFINITELY, EXISTENCE, and OCCURRING. Eight words appeared within the two middle quartiles of both tests: ENVIRONMENT, DISAPPOINT,

TABLE 5

DIFFICULTY INDEX (PROPORTION PASSING EACH ITEM)

	Multiple Choice	Written test	Difference in Proportion Passing MC higher	Written higher
BELIEVE	(233) .879	(212) .80	.079	
SEPARATE	(165) .623	(144) .543	.080	
SIMILAR	(177) .668	(156) .589	.079	
CATEGORY	(146) .551	(101) .381	.170	
WRITING	(196) .740	(197) .743	.003	
REFERRED	(93) .351	(135) .509		.158
EXISTENCE	(107) .404	(77) .291	.113	
OCCASIONALLY	(144) .543	(104) .392	.151	
PRIVILEGE	(131) .494	(64) .242	.252	
ENVIRONMENT	(162) .611	(110) .415	.196	
OCCURRING	(56) .211	(76) .287		.076
CONSCIENCE	(136) .513	(98) .37	.143	
ACCIDENTALLY	(122) .46	(105) .396	.064	
PRECEDE	(127) .479	(103) .389	.090	
RECEIVING	(132) .498	(127) .479	.019	
DISAPPOINT	(147) .555	(125) .472	.083	
IRRELEVANT	(144) .543	(86) .325	.218	
MISSPELLED	(119) .449	(128) .483		.034
UNNECESSARY	(200) .755	(106) .40	.355	
DEFINITELY	(117) .442	(72) .272	.170	
	\10.769	\8.779	\2.265	\.268
Mean of each	.538	.439	.133	.089

TABLE 6

ORDER OF EASE-TO-DIFFICULTY

Multiple Choice		Written Test	
.879	BELIEVE	.80	BELIEVE
.755	UNNECESSARY	.743	WRITING
.74	WRITING	.589	SIMILAR
.668	SIMILAR	.543	SEPARATE
.623	SEPARATE	.509	REFERRED
.611	ENVIRONMENT	.483	MISSPELLED
.555	DISAPPOINT	.479	RECEIVING
.551	CATEGORY	.472	DISAPPOINT
.543	OCCASIONALLY	.415	ENVIRONMENT
.543	IRRELEVANT	.40	UNNECESSARY
.513	CONSCIENCE	.396	ACCIDENTALLY
.498	RECEIVING	.392	OCCASIONALLY
.494	PRIVILEGE	.389	PRECEDE
.479	PRECEDE	.381	CATEGORY
.46	ACCIDENTALLY	.37	CONSCIENCE
.449	MISSPELLED	.325	IRRELEVANT
.442	DEFINITELY	.291	EXISTENCE
.404	EXISTENCE	.287	OCCURRING
.351	REFERRED	.272	DEFINITELY
.211	OCCURRING	.242	PRIVILEGE

CATEGORY, OCCASIONALLY, CONSCIENCE, RECEIVING, PRECEDE, and ACCIDENTALLY.

But on five words, notable discrepancies appeared:

UNNECESSARY - MC index .755 (#2), written index .400 (#10)

IRRELEVANT - MC index .543 (#9-10), written index .325 (#16)

PRIVILEGE - MC index .494 (#13), written index .242 (#20)

MISSPELLED - MC index .449 (#16), written index .483 (#6)

REFERRED - MC index .351 (#19), written index .509 (#5)

7. The discrimination index for each item of each test was calculated (Table 7), using the following formula:

High group proportion right - low group proportion right = discrimination

As previously explained (with Descriptive Item Analysis), the population taking each test was subdivided into high and low criterion groups on the basis of mean scores:

Multiple-choice mean 10.77: 140 scored above mean (11 +)
125 scored below mean (10 -)

Written test mean 8.78: 129 scored above mean (9 +)
136 scored below mean (8 -)

For each multiple choice item the discrimination index was thus calculated using the following specific formula:

$$\frac{\text{High group number right}}{140} - \frac{\text{Low group number right}}{125} = \text{discrimination}$$

The following multiple-choice items were strongest in discriminating ability: ENVIRONMENT .44; PRECEDE .41; ACCIDENTALLY .40; SEPARATE, RECEIVING, .38; WRITING, MISSPELLED, DEFINITELY, .36.

The following multiple-choice items were moderate in discriminating ability: CATEGORY .34; PRIVILEGE .31; REFERRED .30; DISAPPOINT .28;

TABLE 7

DISCRIMINATION INDEX TABLE

	MC High Group Right (N=140)	MC Low Group Right (N=125)	Multiple-Choice Discrimination	Wr High Group Right (N=129)	Wr Low Group Right (N=136)	Written Test Discrimination
BELIEVE	136	97	.19	120	92	.25
SEPARATE	112	53	.38	102	42	.48
SIMILAR	109	68	.24	102	54	.39
CATEGORY	100	46	.34	74	27	.37
WRITING	127	69	.36	123	74	.41
REFERRED	69	24	.30	98	37	.49
EXISTENCE	73	34	.25	65	12	.41
OCCASIONALLY	91	53	.23	76	28	.38
PRIVILEGE	90	41	.31	56	8	.37
ENVIRONMENT	115	47	.44	90	20	.55
OCCURRING	42	14	.19	57	19	.30
CONSCIENCE	90	46	.27	79	19	.47
ACCIDENTALLY	91	31	.40	93	12	.63
PRECEDE	94	33	.41	78	25	.42
RECEIVING	90	32	.38	88	39	.39
DISAPPOINT	96	51	.28	86	39	.38
IRRELEVANT	90	54	.21	67	19	.38
MISSPELLED	87	32	.36	78	50	.23
UNNECESSARY	123	77	.26	85	21	.51
DEFINITELY	86	31	.36	65	7	.45

CONSCIENCE .27; UNNECESSARY .26; EXISTENCE .25.

The following multiple-choice items were weakest in discriminating ability: SIMILAR .24; OCCASIONALLY .23; IRRELEVANT .21; BELIEVE, OCCURRING, .19.

For each written test item the discrimination index was calculated using the following specific formula:

$$\frac{\text{High group number right}}{129} - \frac{\text{Low group number right}}{136} = \text{discrimination}$$

Seventeen of the twenty written items were strong in discrimination of abilities, with three particularly strong indices: ACCIDENTALLY .63; ENVIRONMENT .55; UNNECESSARY .51; REFERRED .49; SEPARATE .48; CONSCIENCE .47; DEFINITELY .45; PRECEDE .42; WRITING, EXISTENCE, .41; SIMILAR, RECEIVING, .39; OCCASIONALLY, DISAPPOINT, IRRELEVANT, .38; CATEGORY, PRIVILEGE, .37. Two were moderate: OCCURRING .30; BELIEVE .25.

Weakest in discrimination of abilities was MISSPELLED .23.

A general discrimination ratio for each test was calculated by totaling the discrimination indeces of its twenty questions and finding their mean.

The multiple-choice test's general discrimination index was .308 (satisfactory but not strong).

The written test had a strong general discrimination index of .413. Of the two tests, it could be said that the written form is better able to discriminate between high- and low-ability students.

8. The null hypothesis was tested, using independent t-test, two-tailed analysis. As explained with Procedures, above, the 0.01 level of significance was used, to minimize the chance of a Type I error and loss of the advantages of multiple-choice testing.

   As seen in Table 8, at this level the null hypothesis must be rejected; there is a significant difference between the mean scores of the two groups.

9. A scatter diagram was plotted (Table 9), to analyze correlation between students' multiple-choice and written scores. A positive linear correlation is seen.

10. The strength of this linear relationship was calculated, using Pearson's formula for coefficient of linear correlation:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\ \sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

in which n = number of students taking the pair of tests (265); x = individual scores on multiple-choice test; and y = individual scores on written test.

A correlation coefficient of .779 was reached, exceeding the critical table value of r (.160), and indicating a definite positive correlation between the two sets of scores. (Statistical computations are included in Appendix D.)

This is reminiscent of the following comment by Wood:

> One frequently encounters the opinion that the serious weakness of objective tests is that they tap "mere recognition".... Several early experiments bear on this question. When the same content was tested by multiple-choice tests and by the completion form in which the subject must recall and then supply missing material, scores were indeed higher for the recognition type. But the significant finding was that scores for the two forms were almost perfectly correlated....Thus the correlation between

TABLE 8

STATISTICAL SUMMARY OF EXPERIMENTAL RESULTS

	N	Mean	Standard Deviation	Variance (s^2)
Multiple-choice test	265	10.77	3.65	13.34
Written test	265	8.78	4.89	23.88

Level of significance = 0.01

Critical t table value = 2.576

Calculated t value = 5.30

HYPOTHESIS $H_o : \bar{X}_{mc} = \bar{X}_w$

Students' mean score of words spelled correctly on the multiple-choice test will equal students' mean score of words spelled correctly on the written test.

(Alternate hypothesis: $H_a : \bar{X}_{mc} \neq \bar{X}_w$)

CONCLUSION Because the calculated value of t (5.30) exceeds the critical value of t ($\pm$ 2.576), the null hypothesis must be rejected. At the 0.01 level of significance, there is a significant difference between the mean scores of the two groups.

Note: Statistical computations are included in Appendix D.

TABLE 9

SCATTER DIAGRAM OF MULTIPLE CHOICE AND WRITTEN TEST SCORES' CORRELATION

Written	Multiple Choice 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Written Frequency Total
20																						
19																		1	1	2	1	5
18																			2	1		3
17																4		4	2			10
16										1			1	1	1	2	1	2				9
15														4	4	2	1		1			12
14												1	2	3	3	6	1					16
13										1		3	2	6	3	4						19
12									1				3	2	3	1	1	2				13
11												2	2	1	2	2		1				10
10								1		2		3	5	2	1							14
9							2	1	2	2	5	3	1	1		1						18
8								3	5	2	2	3	3									18
7						2	1	2	1	5	1	1	3	1	2		1					20
6						3	2	2	2	5	1	6	1									22
5						2		6	3	2	4			2								19
4						1	5	1	2	2	1	4	2									18
3			1		2	1		3	2	2	1											12
2						4		1	1	3	2											11
1		1			1	1		1	2	1			1									8
0						1	2	4					1									8
Multiple-Choice Freq. Total	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
		1	1		2	15	12	25	21	28	17	26	27	23	19	22	5	10	6	3	1	265

highly reliable tests eliciting recall and those demanding only recognition is so high that for all practical purposes the scores are interchangeable.[23]

11. The estimated regression line was then calculated by the least squares method, using the following formula:

$$y \text{ predicted} = m_y + b\,(x - m_x)$$

$$b = \frac{n\,(\Sigma xy) - (\Sigma x)\,(\Sigma y)}{n\,(\Sigma x^2) - (\Sigma x)^2}$$

From these formulas, a specific formula was derived for predicting a student's score on the written spelling test (given his score on the multiple-choice test):

$$y(\text{written test score}), \text{ predicted} = 1.047\,x - 2.496$$

(x = his multiple-choice score)

From this specific formula was derived Table 10, Written Test Scores Predicted from Multiple-Choice Test Scores.

Based on the correlation coefficient, these predicted scores could be expected to be accurate 77.9% of the time.

The right column of Table 10 indicates the predicted written score's percentage of the multiple-choice score, ranging from 21.5% of a 3 multiple-choice score to 92.22% of a 20 multiple-choice score.

[23]Dorothy Adkins Wood, Test Construction: Development and Interpretation of Achievement Tests (Columbus, Ohio: Charles E. Merrill Books, Inc., 1961), p. 94.

TABLE 10

WRITTEN TEST SCORES PREDICTED FROM MULTIPLE-CHOICE TEST SCORES

(Specific formula derived from estimated regression line, calculated by least squares method, to predict written test score -- given x, a specific multiple-choice test score: $y = 1.047x - 2.496$

Multiple-choice score	Predicted written score	Percent of MC score
0	(-2.496)	
1	(-1.449)	
2	(- .402)	
3	.645	21.5
4	1.692	42.3
5	2.739	54.78
6	3.786	63.1
7	4.833	69.04
8	5.88	73.5
9	6.927	76.97
10	7.974	79.74
11	9.021	82.01
12	10.069	83.91
13	11.115	85.5
14	12.162	86.87
15	13.209	88.06
16	14.256	89.1
17	15.303	90.01
18	16.35	90.83
19	17.397	91.56
20	18.444	92.22

DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

The primary conclusion of the present study is that our students are indeed weak in spelling skills, and the community college should recognize an obligation to help develop those skills. Of a possible 20, the mean score on the multiple-choice test was 10.77; on the written test, 8.78; and of words spelled correctly on both, the mean was 7.07.

The number of words a student spells correctly on both tests may be assumed to reflect his "true" knowledge of these spelling words -- the way he would spell these words in a non-test or composition situation. Since the number spelled correctly on both tests was much lower than the individual test scores, this would seem to reflect students' lack of "true" knowledge of these spelling words, and a greater probability of guessing at correct spelling.

The Descriptive Item Analysis (Table 4, pp.21-24) compared responses on both multiple-choice and written tests -- with each subdivided into high and low groups (above or below mean). Several features are of note:

On the written test, the low-score group was much more "creative" than the high-score group in producing more widely variant spellings (Other Responses, tabulated in column six and listed in Appendix C). Of 1412 "other responses," 283 (20%) came from those scoring above the mean; 1129 (80%) came from those scoring below the mean.

On both tests, the low-score groups were far more inclined to leave questions unanswered (despite instructions: "Please answer <u>every</u> question; if you aren't sure, guess.") On the multiple-choice and written tests, respectively, the high-score groups left 14 and 4

questions unanswered (total 18), vs. the low-score groups left 19 and 157 questions unanswered (total 176).

Some errors seemed more likely on the multiple-choice test: Twelve incorrect spellings which had not occurred at all on the written test were selected a total of 85 times on the multiple-choice test: SEPARETE (7), SEPPERATE (1), CATAGARY (4), WRITTEING (1), RITEING (1), EXISTANSE (7), CONSCIUS (5), PROCEDE (22), MISSPELED (19), UNNESARY (1), UNNESISARY (9), UNECISSARY (8). And 29 incorrect spellings appeared 6 or fewer times on the written test (a total of 80 misspellings), but on the multiple-choice test, these same 29 misspellings were selected 284 times (e.g. the incorrect spelling OCASSIONALY was spontaneously produced only 6 times, but selected 35 times; MISS-SPELLED was spontaneously produced only 6 times, but selected 30 times; UNNECESARY was spontaneously produced only 5 times, but selected 46 times).

This discrepancy could have several possible explanations: A student who guessed the wrong spelling on the written test -- or who thought he knew, but was incorrect -- is limited on the written test to selecting one of five options. His incorrect spontaneous spelling may not be one of the five options, or he may not remember his previous answer. He must make another selection.

Or he may have spelled the word correctly on the written test, but then simply have misread or mistaken the multiple-choice option he chose. Or perhaps on the written test he was able to think of only a single spelling (right or wrong), and so didn't have to make any decision. Confronting the five options, however, may have caused uncertainty and a wrong choice. In any event, several misspellings which were rarely

or never thought of during the written test seemed plausible enough to select on the multiple-choice test.

The written test (which has been assumed to reflect more nearly the spontaneous spelling a student would use in writing a composition) proved to be more difficult than the multiple-choice -- with a lower mean score (8.78) and general difficulty ratio of .439 (mean proportion passing per item), in contrast to the multiple-choice test's higher mean score (10.77) and higher difficulty ratio (mean proportion passing per item) of .538. On seventeen of the twenty words tested, a higher proportion passed on the multiple-choice test than on the written test.

This implies that the average student's score on a multiple-choice test of given words would be "inflated" -- better than his spontaneous spelling ability. This "inflation" was borne out by the t-test: Even at the 0.01 level, the difference between mean scores of the two tests was significant, and the null hypothesis -- that these mean scores would <u>not</u> vary significantly -- had to be rejected.

The multiple-choice spelling test, however, with its advantages of greater speed, accuracy, and objectivity, is not without value. The two sets of scores were shown to have a positive linear correlation coefficient of .779. Keeping this proportion in mind, the multiple-choice test could be used in several recommended ways within a spelling program.

1. Perhaps the most accurate testing method for predicting a student's spontaneous spelling in a non-test or composition situation would be to administer the same written test on two different occasions -- or to administer both multiple-choice and written tests (with same words) on the same occasion -- and count the number he spells correctly <u>twice</u>.

If such double testing is not feasible, the .779 correlation coefficient suggests that a reasonably accurate prediction could be achieved with the multiple-choice test alone. These score conversions are shown in Table 10, above.

2. Results of the present study could be used to develop a nultiple-choice spelling contingent of a placement test -- to provide a rapid estimate of the applicant's spelling ability. For this purpose, words with strongest discrimination indices should be used -- those best able to discriminate between high- and low-ability students. Some distracters on the multiple-choice test were of questionable value, if chosen by few or none, or by more in the high group than the low group. The discrimination ability of the multiple-choice questions might be strengthened by replacing poor distracters with incorrect spellings produced most often on the written test. For this purpose, a revision of distracters is recommended in Table 11.
3. Alternating written and multiple-choice tests throughout a spelling program, with ongoing analysis of results, could thus provide an optimum combination -- maintaining insofar as possible the advantages of each method, while providing built-in correction and compensation for each other's disadvantages.

   In this manner it is probable that the multiple-choice test's discrimination ability, reliability, and correlation coefficient could all gradually be improved -- perhaps to the point that testing the null hypothesis would eventually reflect "no significant difference."
4. Greatest spelling mastery of a given group of words could no doubt be achieved by repeated work and testing, and perhaps by alternating

TABLE 11

RECOMMENDED MULTIPLE-CHOICE DISTRACTERS (FOR FUTURE USE)
(with number of spontaneous misspellings in the present study)

BELIEVE: BELEAVE 6, BELEIVE 16, BELIEF 2, BELIVE 17

SEPARATE: SEPARTE 8, SEPERATE 75, SEPPARATE 2, SEPRATE 11

SIMILAR: SIMALAR 6, SIMILIAR 35, SIMLAR 5, SIMULAR 24

CATEGORY: CATAGORIE 4, CATAGORY 80, CATEGORIE 4, CATOGORY 8

WRITING: WRITEING 4, WRITEN 1, WRITTEN 6, WRITTING 50

REFERRED: REFER 8, REFERED 73, REFFERED 14, REFFERRED 2

EXISTENCE: EXISTANCE 84, EXSISTANCE 9, EXSISTENCE 4, EXSITENCE 7

OCCASIONALLY: OCASSIONALLY 14, OCCASIONALY 8, OCCASIONLY 10, OCCASSIONALLY 27

PRIVILEGE: PRIVELEDGE 10, PRIVELEGE 10, PRIVILAGE 25, PRIVILEDGE 38

ENVIRONMENT: ENVIORMENT 15, ENVIORNMENT 13, ENVIROMENT 39, INVIROMENT 8

OCCURRING: ACCURING 4, OCCURING 124, OCURING 5, OCURRING 9

CONSCIENCE: CONCIENCE 10, CONCIOUS 10, CONSCIOUS 45, CONSIOUS 5

ACCIDENTALLY: ACCIDENTALY 19, ACCIDENTIALLY 4, ACCIDENTLEY 2, ADDICENTLY 81

PRECEDE: PRECEED 86, PRECEEDE 7, PRECIDE 4, PRESEDE 3

RECEIVING: RECEIVEING 3, RECEVING 4, RECIEVING 89, RECIVING 5

DISAPPOINT: DISAPOINT 41, DISPOINT 9, DISSAPOINT 44, DISSAPPOINT 11

IRRELEVANT: IRRELAVENT 11, IRRELEVENT 16, IRREVELANT 9, IRREVELENT 9

MISSPELLED: MISPELL 15, MISPELLED 49, MISSPELL 44, MISS-SPELLED 6

UNNECESSARY: UNECESSARY 14, UNNECCESARY 6, UNNECCESSARY 16, UNNESSARY 29

DEFINITELY: DEFINATELY 31, DEFINATLY 18, DEFINETLY 16, DEFINITLY 10

Note: Underlined distracters were used in the present study, and are recommended for reuse.

written and multiple-choice tests. An optimum program might include the following processes:

1. A written pretest with traditional oral dictation and context sentences
2. Tabulation of responses, analysis of students' errors, and instruction (ideally, on an individual or small-group basis) in avoiding those errors.
3. Each student would make a spelling log, including analysis of his own errors, and mnemonics to avoid those errors and achieve correct spellings.
4. A multiple-choice practice test, with distracters being the most common misspellings on the written pretest.
5. Further tabulation and analysis of responses.
6. Each student continues his log by entering these errors, to compare results: same errors vs. new errors vs. improvement.
7. A written post-test.
8. Statistical analyses, comparing
   Written pretest and written post-test
   Written pretest and multiple-choice test
   Multiple-choice test and written post-test
9. Those lacking mastery could be retested by multiple-choice and written tests, alternately (with log entries and analysis after each) until desired mastery is achieved.

(If time requires a smaller-scale spelling program, it should maintain as many of the above processes and features as possible.)

A program of this sort would certainly be worthy of further research.

A multiple-choice spelling test is presently a less accurate reflection of the student's spontaneous spelling ability than a written spelling test. But in the ways proposed, the multiple-choice spelling test can proved to be of increasing value, and can thus make a worthwhile contribution to improving the writing and spelling skills of the community college student.

APPENDICES

APPENDIX A

ANSWER STRIP FOR
WRITTEN SPELLING TEST

(The written test is to be administered before the multiple-choice test.)

To insure maximum uniformity, the oral dictation is presented on a cassette tape ($6\frac{1}{2}$ minutes).

The instructions are read; then each word is pronounced, used in a sentence, and repeated.

There are the same words and sentences on the multiple-choice test (on the following page).

Name _____

Soc.Sec.# _____

Section: English 101._____

Your class is participating in a spelling experiment, in two parts.

This first part is a written spelling test: you're asked to write or print 20 words.

On each, I'll pronounce the word, then use it in a sentence, then repeat it.

Start immediately to write or print the word.

Please answer <u>every</u> question; if you aren't sure, guess.

Make every letter <u>clear</u>.

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

12.

13.

14.

15.

16.

17.

18.

19.

20.

Now turn in this answer sheet, and you'll be given the multiple-choice part of the test.

APPENDIX B

Name _____ 45

Soc.Sec.# _____

MULTIPLE-CHOICE SPELLING TEST

Section: English 101._____

Circle one answer per item.

Please answer every question; if you aren't sure, guess.

1. I -- you.
   A- BELEIF, B- BELIEF, C- BELIEVE, D- BELEIVE, E- BELEEVE

2. They have -- offices.
   A- SEPERETE, B- SEPARETE, C- SEPERATE, D- SEPARATE, E- SEPPERATE

3. Their ideas are --.
   A- SIMALAR, B- SIMILAR, C- SIMULAR, D- SIMALIAR, E- SIMILIAR

4. He is in amateur --.
   A- CATAGORY, B- CATEGORY, C- CATOGORY, D- CATAGARY, E- CATEGARY

5. She is -- a book.
   A- WRITING, B- WRITEING, C- WRITTING, D- WRITTEING, E- RITEING

6. He -- to the dictionary.
   A- REFERED, B- REFFERED, C- REFFERD, D- REFERRED, E- REFFERRED

7. The eagle's -- is threatened.
   A- EXISTANCE, B- EXISTANSE, C- EXISTENCE, D- EXISTENSE, E- EXSISTENCE

8. He -- smokes cigars.
   A- OCASIONALY, B- OCASSIONALY, C- OCASSIONALLY, D- OCCASIONALLY, E- OCCASSIONALLY

9. Voting is a --.
   A- PRIVILEDGE, B- PRIVELEDGE, C- PRIVILEGE, D- PRIVELEGE, E- PRIVILAGE

10. Don't pollute the --.
    A- ENVIORNMENT, B- ENVIRONMENT, C- ENVIORMENT, D- ENVIRAMENT, E- ENVIORAMENT

11. A storm is --.
    A- OCURING, B- OCCURING, C- OCCERING, D- OCURRING, E- OCCURRING

12. She has a guilty --.
    A- CONSCIUS, B- CONSCIENCE, C- CONSCIOUS, D- CONSCEINCE, E- CONSCIENSE

13. He -- broke the glass.
    A- ACCIDENTLY, B- ACCIDENTALY, C- ACCIDENTALLY, D- ACCIDENTILY, E- ACCIDENTILLY

14. June will -- July.
    A- PRECEDE, B- PRECEED, C- PRECEEDE, D- PROCEDE, E- PROCEEDE

15. She is -- many letters.
    A- RECIEVING, B- RECIEVEING, C- RECEIVING, D- RECEIVEING, E- RECEVING

16. I'm sorry to -- you.
    A- DISAPPOINT, B- DISAPOINT, C- DISSAPOINT, D- DISSAPPOINT, E- DISSAPPOINTE

17. That information is --.
    A- IRRELEVENT, B- IRRELEVANT, C- IRRELAVENT, D- IRREVELANT, E- IRREVELENT

18. He -- four words.
    A- MISS-SPELLED, B- MISPELED, C- MISSPELED, D- MISPELLED, E- MISSPELLED

19. This trip is --.
    A- UNNESARY, B- UNNESISARY, C- UNECISSARY, D- UNNECESARY, E- UNNECESSARY

20. You are -- right.
    A- DEFINATLY, B- DEFINATELY, C- DEFINETELY, D- DEFINITELY, E- DEFINITLY

APPENDIX C

FREQUENCY DISTRIBUTIONS OF WRITTEN TEST RESPONSES

6	BELEAVE		ACEPTIVE		SALMBER		CAGAORY
	BELEEVE *		SACTRIC		SAMARLIAR		CAGIORGE
	BELEIF *		SAPARATE		SAMILIR		CAGORIES
16	BELEIVE *		SAPERATE		SEAMILAR		CALGORIA
2	BELEVE		SAPRATE		SEMBLER		CAPITORY
	BELEVER		SARPARTATE	3	SEMILAR		CARGORLY
	BELEVVD		SEFRTE		SEMLER		CATACOLE
2	BELIEF *		SELPER		SILMILIAR	0	CATAGARY *
212	<u>BELIEVE</u>		SEOBYRED	6	SIMALAR *		CATAGOLY
17	BELIVE		SEPAERATE	2	SIMALIAR *	4	CATAGORIE
	BELIVEI	144	<u>SEPARATE</u>		SIMARLE	80	CATAGORY *
	BELIEVING		SEPARATLY		SIMARLIAR		CATAGOTARY
	BELVIED		SEPARET	2	SIMER		CATAGOTERY
	BIELIEVE	0	SEPARETE *		SIMERLAR		CATAORY
	BLEIVE	8	SEPARTE		SIMERLY		CATDERGOR
			SEPATOR		SIMILAIR		CATEGAL
		75	SEPERATE *	156	<u>SIMILAR</u>		CATEGARY *
			SEPEREATE		SIMILARY		CATEGERY
		0	SEPERETE *		SIMILER		CATEGOARY
		2	SEPPARATE	35	SIMILIAR *		CATEGOREY
		0	SEPPERATE *	2	SIMILLAR	4	CATEGORIE
		11	SEPRATE		SIMILUR	101	<u>CATEGORY</u>
			SEPRETE		SIMIRLAR		CATEOGORY
			SEPTRATE	5	SIMLAR		CATERGIOUS
			SPEATE		SIMLIR		CATERGOLY
		2	SPERATE		SIMMILAR	4	CATERGORY
			SPRAT		SIMPLAR		CATEROGORY
			SPREATE	24	SIMULAR *	2	CATEROGY
			SPREATED		SIMULER		CATGOREY
			SUPRATE		SLEMBER		CATGROTARY
					SMERLAE	3	CATIGORY
					SMUILAR		CATOFGO
					SUMMIER	2	CATOGARY
							CATOGATE
							CATOGLORE
							CATOGOLY
						2	CATOGORE
						2	CATOGORIE
						8	CATOGORY *
							CATOGOY
							CATOGRARY
							CATORCOREY
							CATORGIE
							CATORGOLY
							CATORGORE
							CATORGORIE
						2	CATORGORY

(continued)

Notes: Correct answers are underlined.
* denotes a multiple-choice distracter.
Each response was received once unless otherwise indicated.

Appendix C, continued
Frequency Distributions of Written Test Responses

2 CATORGY
CATOROGY
CATORRY
CATORY
CETORGOY
CORGAD
CUTAGGORRY
KATTEGORY
VOCABLARY

RINDING
RIRETTING
0 RITEING *
WIRTTEN
WRIRTING
2 WRITE
4 WRITEING *
WRITEN
197 WRITING
0 WRITTEING *
6 WRITTEN
50 WRITTING *

PREFER
REBERED
3 REFEARD
2 REFEARED
REFEARRED
8 REFER
REFERD
73 REFERED *
REFEREND
REFERIED
135 REFERRED
2 REFFERD *
REFFERENCE
14 REFFERED *
2 REFFERRED *
REFORD
REFORDENCE
2 REFORED
REFRED
3 REFURED
REFURRED
REPRERREED
RERFERRED
REROAR
REVAN

ASSITANCE
ASSITENCE
2 ASSISTANCE
ATZESTION
ECSISTENCE
EGISTANIST
ENDISTION
ENVISENCE
ERZINT
EXANTANCE
EXAUST
EXCESTANCE
EXCINTANT
EXCISCT
EXCISES
2 EXCISTANCE
EXCISTENCE
EXCISTEND
EXCISTENS
EXCITANT
EXCITEN
EXCITENCE
EXCITES
EXCITION
EXDUSTON
EXECISTANCE
EXHISTANCE
EXICANCE
EXICSTENCE
EXIENTENCE
EXIESTANTS
EXISISTANCE
EXISISTION
84 EXISTANCE *
0 EXISTANSE *
EXISTANST
EXISTANTT
77 EXISTENCE
EXISTENING
2 EXISTENSE *
EXISTENT
EXISTIANCE
EXITAL
2 EXITANCE
EXITENCE
EXITENSE
EXITICED
EXITING
EXITSTANCE
EXSIENCE

EXSIENTANCE
EXSINTANCE
9 EXSISTANCE
4 EXSISTENCE *
EXSISTENSE
EXSISTING
2 EXSISTION
EXSITANCE
EXSITANTES
7 EXSITENCE
EXSTINCT
2 EXTANCE
EXTENANCE
EXTENSION
EXTENTIST
EXTIANCE
EXTISIED
EXZENGT

(continued)

Appendix C, continued
Frequency Distributions of Written Test Responses

ACCASIONLY
ACCATIONALLY
ACCOLLING
OCAISIONLY
OCALAGLE
OCANANCY
OCANCIONALY
OCASIOLLY
OCASION
4 OCASIONALLY
3 OCASIONALY *
14 OCASSIONALLY *
6 OCASSIONALY *
2 OCASSIONLY
OCASSLION
OCATEGELY
2 OCATIONALLY
2 OCATIONALY
OCATLET
2 OCCAISIONALLY
OCCANCIOLLY
OCCANICALLY
OCCANILE
OCCANIONALLY
OCCANIONLY
OCCASHIONALY
OCCASIANNALLY
OCCASIOALLY
OCCASIONAL
104 OCCASIONALLY
8 OCCASIONALY
OCCASIONLLY
10 OCCASIONLY
OCCASSALLY
OCCASSIONAL
27 OCCASSIONALLY *
6 OCCASSIONALY
OCCASSIONILY
OCCASSIONLLY
3 OCCASSIONLY
OCCATION
2 OCCATIONALLY
OCCATIONALY
3 OCCATIONLY
OCCCULARY
OCCIASIONLY
OCCIONALLY
OCCIONTALY
OCCLUGER
OCCOLALANT
(continued)

OCCONALLY
OCCONAT
OCCOULLY
OCCUASIONALLY
OCCUASIONALY
OCCUASLY
OCCUATIONALLY
OCCULAN
OCCULATION
OCCULY
OCCUSEANALY
OCCUSENLY
OCIALLEY
OCURRELY
OCUSSALEY
OKASHLY
OKYTIONALY
OSCASIONALLY

PERVALEDGE
PRAVELAGE
PRAVILAGE
PREAGE
PREGELI
PREILIZE
PRELIVIAGE
PREQTER
PREPRIAL
PREVALEGE
PREVALENG
2 PREVALIGE
PREVELATE
2 PREVELEGE
PREVELIAGE
PREVELIGE
PREVELISH
PREVELITIES
PREVELLAGE
PREVENGE
PREVEOULY
PREVIALEGE
PREVIBLE
7 PREVILAGE
PREVILEDGE
PREVILEG
5 PREVILEGE
PREVILIAGE
2 PREVILIGE
PREVILLEG
PREVILLEGE
2 PREVIOUSLY
2 PREVLEDGE
PREVLISO
PREVORLIGY
PREWLAGE
PRIELEVESE
PRIETPRV
PRIEVELIES
PRILAGE
PRILEGE
PRILIEVESY
PRIPREGART
PRIVAGE
3 PRIVALAGE
4 PRIVALEDGE
6 PRIVALEGE
PRIVALIDGE
PRIVALIGE
PRIVEDGE
(continued)

PRIVEGE
2 PRIVELAGE
10 PRIVELEDGE *
PRIVELEGDE
10 PRIVELEGE *
2 PRIVELGE
PRIVELY
25 PRIVILAGE *
PRIVILAVY
PRIVILDGE
38 PRIVILEDGE
64 PRIVILEGE
PRIVILEGIE
PRIVILIAGE
PRIVILIDGE
4 PRIVILIGE
PRIVILLEDGE
PRIVILY
2 PRIVLAGE
PRIVLEAGE
5 PRIVLEDGE
3 PRIVLEGE
PRIVLIGE
PRIVLLE

Appendix C, continued
Frequency Distributions of Written Test Responses

ENIORMENT
ENIOROMENT
ENIROMENT
ENIVIORMENT
2 ENIVORNMENT
ENVARIMENT
ENVIORAMENT *
15 ENVIORMENT *
13 ENVIORNMENT *
ENVIOROMENT
ENVIORONMENT
ENVIORTMENT
2 ENVIOURMENT
ENVIRAMENT *
2 ENVIREMENT
ENVIRERMOMENT
ENVIRIMENT
3 ENVIRMENT
2 ENVIRNMENT
ENVIRNOMENT
39 ENVIROMENT
ENVIROMENTT
110 ENVIRONMENT
ENVIRTMENT
ENVOIMENT
ENVOIRMENT
ENVOLIMENT
ENVOREMENT
ENVORIALMENT
ENVORIMENT
ENVORIOMENT
2 ENVORMENT
ENVORONMENT
ENVRIOMENT
ENVRIONMENT
ENVRONMENT
EVANLE
EVIONLEMENT
2 EVIRONMENT
EVROINMENT
INFONMENT
INVALANENT
INVARNMENT
INVAYERMENT
INVEMMENT
3 INVERMENT
INVERONMENT
INVESTMENT
INVIOMENT
2 INVIORMENT
2 INVIORNMENT
(continued)

INVIREOMENT
INVIRMENT
INVIRNMENT
8 INVIROMENT
6 INVIRONMENT
INVMENT
INVORIMENT
INVORINMENT
INVORMENT
2 INVORNMENT

CANCHANZ
CANSITNES
CENSIOUNS
CHONCH
COENCE
COMSION
CONAGE
CONCEINCE
CONCENCE
CONCENCES
CONCHACE
CONCHANCE
CONCIANCE
(continued)

ACCARING
ACCERRTT
ACCORING
ACCOURING
ACCOURNED
ACCUREING
ACCUREN
ACCURINCE
4 ACCURING
ACCURRING
ACORING
ACORRING
ACURRING
AQUARLING
COURING
OBCURING
OCARING
OCCARRING
OCCERING *
2 OCCOURING
OCCQULING
OCCUR
OCCUREING
OCCURELY
OCCUREN
OCCURENCE
3 OCCURENT
124 OCCURING *
OCCURYING
2 OCCURREN
OCCURRENT
76 OCCURRING
OCURIED
5 OCURING *
9 OCURRING *
OKORIEN
ORCARING
ORLCAN

CONCIEION
10 CONCIENCE
CONCIENSE
CONCIL
CONCINDE
10 CONCIOUS
CONCOSIENCE
CONCOU
CONCOUNCO
CONCUSION
CONGEANTES
CONICENTS
(continued)

CONINESS
CONIOUS
CONISOIOUS
CONJUNCTIVE
CONSANTANCE
CONSCENCE
CONSCENCIOUS
2 CONSCEINCE *
98 CONSCIENCE
CONSCIENOUS
CONSCIENSE *
2 CONSCIENT
CONSCIENTIOUS
CONSCIENTOUS
2 CONSCINCE
CONSCINE
CONSCINT
CONSCION
CONSCIOUNS
45 CONSCIOUS *
0 CONSCIUS *
CONSCNIOUS
CONSECINCE
CONSECIOUS
2 CONSENCE
CONSHENCE
CONSHENSE
CONSHIONES
CONSICENCE
CONSIENCE
CONSINCIOUS
2 CONSION
CONSIONCE
CONSIOUN
5 CONSIOUS
CONSTIONS
CONSUN
CONSUNCE
CONTION
CONTIONCE
CONTIOUS
CONTRUCTION
CONTUSH
CONUSION
COSIENCE
COUNCANCE
COUNIOUS
COUSENT
CUNCH
CUNCHANCE
CUNTOUGIEN
(continued)

Appendix C, continued
Frequency Distributions of Written Test Responses

	ACCADENTLY		PAESEAD		RECEAVING		DECIPOINT
2	ACCDENTLY		PARRED		RECEFTE		DEPOINT
2	ACCEDENTALLY		PECEDE		RECEIAVING		DESAPORT
	ACCEDENTALY	4	PERCEDE		RECEIEVING		DESAPPOINT
	ACCENDENTAL		PERCEDED		RECEIPVING		DESIPPOINT
	ACCENDENTLY		PERCEED		RECEIVABLE		DIOPINT
	ACCENDITLY		PERCEIDE	2	RECEIVE		DISAPOIMENT
	ACCETDENLLY		PERCES		RECEIVED	41	DISAPOINT *
	ACCETENTLY		PERCIEVE	3	RECEIVEING *		DISAPONINT
	ACCIDENTALEY		PERSEE		RECEIVEN	2	DISAPONT
105	ACCIDENTALLY		PERSENSE		RECEIVENG	2	DISAPPIONT
19	ACCIDENTALY *		PRCEDE	127	RECEIVING	125	DISAPPOINT
	ACCIDENTIAL		PREADLEY		RECERIVING		DISAPROVE
4	ACCIDENTIALLY		PRECCED		RECEVE		DISCIPOINT
	ACCIDENTIALY		PRECEADE	2	RECEVEING		DISERPIT
2	ACCIDENTILLY *	3	PRECED	4	RECEVING *		DISIPIONT
	ACCIDENTILY *	103	PRECEDE		RECIEEING		DISOPIONT
2	ACCIDENTLEY		PRECEDED		RECIESDT		DISOPOINT
	ACCIDENTLLY		PRECEDENCE	2	RECIEVEING *		DISOPPONT
81	ACCIDENTLY *		PRECEDENT		RECIEVERING		DISPIONT
	ACCIDENTTALLY	86	PRECEED *	89	RECIEVING *	9	DISPOINT
	ACCIDENTY	7	PRECEEDE *		RECING	3	DISPONT
	ACCIEDENTALLY		PRECEEDED		RECIVE	3	DISPPOINT
	ACCIEDENTLY		PRECEES	3	RECIVEING	44	DISSAPOINT *
	ACCIENDENTALLY		PRECEID		RECIVEN		DISSAPONIT
2	ACCIENDTLY		PRECEIDE		RECIVENT		DISSAPPIONT
	ACCINDENLE		PRECEIED	5	RECIVING	11	DISSAPPOINT *
2	ACCINDENTLY		PRECEIP		RECVEING	0	DISSAPPOINTE
	ACCTANDTLY	7	PRECEIVE		REMOVING		DISSOPOINT
	ACCUDENTALY		PRECEND		RESEAVING		DISSPIONT
	ACDIDENTY		PRECESS		RESIVING		THESPAINT
	ACIDENTINLY	4	PRECIDE		REVECIVEING		THISPONT
	ACTERDENLY	2	PRECIED				
	ACTIDENTLY		PRECIEDE				
	ACTION	3	PRECIEVE				
	ACTUDELLY		PRESCIDE				
	ADCIDENTLY		PRESEAD				
	AKTLAT	3	PRESEDE				
	ANSDENTLY		PRESEID				
	ASSDEFELY		PRESIDE				
	ASSIENDLY		PRESIEVE				
	ATSENDENLY		PRICI				
	AXADENTALLY		PRICY				
	AXETEALLY	4	PROCEED				
	AXIDENTELY		PROCEEDE *				
	BROKE		F..ODUCED				
	EXIOFDENT						
	ICIDENTALLY						

Appendix C, continued
Frequency Distributions of Written Test Responses

AREVELENT
ELEMENT
ELERALVENT
ELEVANT
ELRELVANT
ELVOMENT
ENROLVENT
ENROVENT
ERALAR
2 ERALEVANT
ERALIVATE
ERAVALUTE
ERAVILENT
EREALAVANT
EREALOVENT
EREAVALT
5 ERELAVENT
ERELEMENT
2 ERELEVANT
ERELEVENT
ERELIVANT
ERELVANT
EREVELENT
EREVLENANT
ERIVALENT
ERRALAVENT
ERRALEVENT
ERRALIVANT
ERRALVENT
ERRANTANT
ERRAVALENT
ERRAVEMENT
ERRELAENT
ERRELAVANT
ERRELAVENT
ERRELEVENT
ERREVALENT
ERRILLIVENT
ERREVATE
ERVALANT
EVANATE
EVELERANT
EVELOVENT
EVERALENT
EVERALOVENT
EVERLAVANT
EVERLENT
ILLIAVONT
ILLRALIVENT
INRULAVENT
IRALIVENT

IRELAVANT
IRELAVENT
IRELEVANT
IRELOVENT
IREVELENT
IRRALEVENT
IRRALVENT
IRRAVALLANT
IRRAVENT
IRREALENT
IRREFERRED
IRRELANCE
10 IRRELAVANT
11 IRRELAVENT *
IRRELENT
IRRELEVANAT
86 IRRELEVANT
16 IRRELEVENT *
IRRELEVLANT
4 IRRELIVANT
3 IRRELIVENT
2 IRRELLAVENT
IRRELLEVANT
IRRELLEVENT
IRRELLIVENT
4 IRRELVANT
IRRELVENT
IRREVALANT
3 IRREVALENT
9 IRREVELANT *
9 IRREVELENT *
IRREVERALENT
2 IRREVOLENT
IRREYLAVENT
IRRILAVANT
IRRILAVENT
IRRLENT
IRROLOVENT
IRVENLENT
REELEVANT
RELEVANT
RERELAVENT

MESPELL
MISPPEAL
MISPEEL
MISPELD
4 MISPELED *
15 MISPELL
49 MISPELLED *
MISSELP
MISSESPELL
MISSPALT
0 MISSPELED *
44 MISSPELL
MIS-SPELL
128 MISSPELLED
2 MIS-SPELLED
MISSPELT
MISSPELTED
MISSPLED
MISSSPELL
MISS-SPELL
6 MISS-SPELLED *
MIXSPEAL

ANECESSARY
ONNECCESRAY
UNCENESSARY
2 UNCESSARY
UNDERNESSILE
UNECCECARY
UNECCESSARY
UNECESARILY
14 UNECESSARY
0 UNECISSARY *
UNENSSARY
UNESASARY
UNESCESSARY
UNESESARY
UNESIRET
6 UNESSARY
UNESSASARY
UNESSCERY
2 UNESSECARY
UNESSOCURY

UNESTUR
UNNASSECRY
UNNCESSARY
UNNEASSARY
UNNEASSY
UNNEASUARY
2 UNNESSARY
UNNECCASARY
UNNECCASSARY
UNNECCERAY
UNNECCERERY
6 UNNECCESARY
UNNECCESARRY
16 UNNECCESSARY
UNNECCESSERY
UNNECCESSITARY
UNNECCITARY
5 UNNECESARY *
106 UNNECESSARY
2 UNNECSSARY
UNNESACARY
0 UNNESARY *
UNNESCESARY
UNNESCESSARY
UNNESCRCERY
UNNESERSSIM
UNNESESSARY
0 UNNESISARY *
UNNESRICARY
29 UNNESSARY
UN-NESSARY
2 UNNESSARRY
UNNESSARARY
UNNESSARITY
2 UNNESSASARY
3 UNNESSCARY
UNNESSCERRY
UNNESSEARY
2 UNNESSECARY
UN-NESSECARY
UNNESSELY
UNNESSER
UNNESSERALY
4 UNNESSESARY
UNNESSESSARY
UNNESSIARY
2 UNNESSISARY
UNNESTERSORE
UNSECARRY
UNSUESARY

(continued)

(continued)

Appendix C, continued
Frequency Distributions of Written Test Responses

DAFINATLY
DEAFINANTLY
DEATHLY
DEFAMANTLY
DEFANLY
DEFANTLY
DEFENATELY
DEFENATLY
2 DEFENENTLY
3 DEFENITELY
2 DEFENITLY
DEFENLY
DEFENTANLY
2 DEFENTLY
DEFIANT
DEFIANTLY
3 DEFINALY
2 DEFINATALLY
DEFINATELLY
31 DEFINATELY *
DEFINATLEY
18 DEFINATLY *
DEFINATTEY
3 DEFINELY
DEFINENTLY
5 DEFINETELY *
DEFINETIVELY
DEFINETLEY
16 DEFINETLY
DEFINIATLY
DEFINIENT
DEFINITALY
DEFINITE
DEFINITELLY
72 DEFINITELY
DEFINITLEY
10 DEFINITLY
2 DEFINLY
DEFINNLY
2 DEFINTALLY
DEFINTE
2 DEFINTELY
3 DEFINTLY
DEFFANETLY
DEFFENTILY
5 DEFFENTLY
DEFFILEMT
DEFFINATELY
DEFFINATLY
DEFFINETALLY
DEFFINITELY
DEFITNALLY
DEFREMELY
DELIFONENT
DENEFITELY
DENFINALLY
DENFINATLY
DEPENESSLY
DIATHLY
DIENERY
DIFENTLY
DIFFANTLY
3 DIFFENETLY
5 DIFFENTLY
3 DIFFERENTLY
DIFFERLY
DIFFIANTLY
DIFFICENTLY
DIFFINATLY
DIFFINENTLY
DIFFINETELY
DIFFINETLY
2 DIFFINITELY
DIFFLE
DIFFUCTLY
DIFINATELY
DIFINELY
DIFINITELY
DIFINITLY
DIVELY

(continued)

APPENDIX D

STATISTICAL COMPUTATIONS

T-TEST OF NULL HYPOTHESIS

$$t = \frac{\bar{X}_{mc} - \bar{X}_w}{\sqrt{\frac{s_{mc}^2}{n_{mc}} + \frac{s_w^2}{n_w}}}$$

$$= \frac{10.77 - 8.78}{\sqrt{\frac{13.34}{265} \quad \frac{23.88}{265}}}$$

$$= \frac{1.99}{\sqrt{.05033962 + .09011321}}$$

$$= \frac{1.99}{\sqrt{.14045283}}$$

$$= \frac{1.99}{.375}$$

$t = 5.30$ (critical t table value = 2.576)

$\bar{X}_{mc}$ = multiple-choice mean 10.77

$\bar{X}_w$ = written mean 8.78

s_{mc}^2 = multiple-choice variance 13.34 (standard deviation^2)

s_w^2 = written variance 23.88 (standard deviation^2)

n_{mc} = 265

n_w = 265

PEARSON COEFFICIENT OF CORRELATION

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

n = number of students

x = scores on multiple-choice

y = scores on written test

$$= \frac{265(28957) - (2867)(2331)}{265(34589) - 8219689 \quad 265(26945) - 5433561}$$

$$= \frac{7673605 - 6682977}{946396 \quad 1706864}$$

$$= \frac{990628}{1270973.2}$$

$$= .7794$$

Since the calculated value of r (.779) is greater than the critical table value of r (.160 approximately, at 0.01 level with n = 265, d = 263; interpolated from table values at 0.01 with d = 200, r = .181; with d = 300, r = .148),[24] there is a definite positive correlation between students' scores on multiple-choice and written spelling tests.

[24]F. James Rohlf, and Robert R. Sokal, Statistical Tables (San Francisco, W. H. Freeman and Company, 1969), pp. 224-226.

ESTIMATED REGRESSION LINE, CALCULATED BY LEAST SQUARES METHOD

$$y \text{ (written test score, predicted)} = m_y + b\,(x - m_x)$$

x = multiple-choice test score

$$b = \frac{n\,(\Sigma xy) - (\Sigma x)\,(\Sigma y)}{n\,(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{265\,(28957) - (2867)\,(2331)}{265\,(34589) - 8219689}$$

$$= \frac{7673605 - 6682977}{9166085 - 8219689}$$

$$= \frac{990628}{946396}$$

$$b = 1.047$$

$$y, \text{ predicted} = 8.78 + 1.047\,(x - 10.77)$$

$$= 8.78 + 1.047x - (1.047)\,(10.77)$$

$$= 8.78 + 1.047x - 11.276$$

$$y, \text{ predicted} = 1.047x - 2.496$$

BIBLIOGRAPHY

Anastasi, Anne. Psychological Testing. New York: MacMillan Publishing Co., Inc., 1968.

Braddock, Richard, Richard Lloyd-Jones, and Lowell Schoer. Research in Written Composition. Champaign, Illinois: National Council of Teachers of English, 1963.

Buros, Oscar Krisen, ed. "Spelling," The Seventh Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1972, 1:510-514.

Campbell, Donald T., and Julian C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally College Publishing Company, 1963.

Choppin, B. "Guessing the Answer on Objective Tests," British Journal of Educational Psychology, 45:206-213 (June, 1975).

Diederich, Paul B. "Short-Cut Statistics for Teacher-Made Tests," Statistics and Measurement in the Classroom, ed. Carl F. Hereford, Luiz F. S. Natalicio, and Susan J. McFarland. Duguque, Iowa: Kendall/Hunt Publishing Company, 1969.

Dressel, Paul L. Handbook of Academic Evaluation: Assessing Institutional Effectiveness, Student Progress, and Professional Performance for Decision Making in Higher Education. San Francisco: Jossey-Bass Publishers, 1976.

Green, John A. Teacher-Made Tests. New York: Harper & Row, Publishers, 1963.

Higgins, John A. "Remedial Students' Needs vs. Emphases in Text-Workbooks," College Composition and Communication, 24:188-192 (May, 1973).

Kelso, Genell. Spelling Improvement Units Taught by Classroom Drill vs. Self-Instructional Package: Their Development and Comparative Evaluation. Fort Lauderdale, Florida: Nova University, September 1977 (ERIC, number not yet available).

Lord, Frederic M. "Optimal Number of Choices Per Item--A Comparison of Four Approaches," Journal of Educational Measurement, 14:33-38 (Spring, 1977).

Palmer, Osmond E. "Evaluation of Communication Skills," Evaluation in Higher Education, Paul L. Dressel and others, ed. Herold C. Hunt. Boston: Houghton Mifflin Company, 1961.

Pollock, Thomas Clark, and William D. Baker. The University Spelling Book. Englewood Cliffs, New Jersey, 1955. Quoted in John C. Hodges and Mary E. Whitten. Harbrace College Handbook, New York: Harcourt Brace Jovanovich, Inc., 1972.

Rohlf, F. James, and Robert R. Sokal. Statistical Tables. San Francisco: W. H. Freeman and Company, 1969.

Saupe, Joe L. "Technical Considerations in Measurement," Evaluation in Higher Education, Paul L. Dressel and others, ed. Herold C. Hunt. Boston: Houghton Mifflin Company, 1961.

Shaughnessy, Mina P. Errors and Expectations: A Guide for the Teacher of Basic Writing. New York: Oxford University Press, 1977.

Wick, John W. Educational Measurement: Where Are We Going and How Will We Know When We Get There? Columbus, Ohio: Charles E. Merrill Publishing Company, 1973.

Wood, Dorothy Adkins. Test Construction: Development and Interpretation of Achievement Tests. Columbus, Ohio: Charles E. Merrill Books, Inc., 1961.